

ON CONSTANT-MULTIPLE-FREE SETS CONTAINED IN A RANDOM SET OF INTEGERS

SANG JUNE LEE

ABSTRACT. For a rational number $r > 1$, a set A of positive integers is called an r -multiple-free set if A does not contain any solution of the equation $rx = y$. The extremal problem on estimating the maximum possible size of r -multiple-free sets contained in $[n] := \{1, 2, \dots, n\}$ has been studied for its own interest in combinatorial number theory and for application to coding theory. Let a, b be positive integers such that $a < b$ and the greatest common divisor of a and b is 1. Wakeham and Wood showed that the maximum size of (b/a) -multiple-free sets contained in $[n]$ is $\frac{b}{b+1}n + O(\log n)$.

In this note we generalize this result as follows. For a real number $p \in (0, 1)$, let $[n]_p$ be a set of integers obtained by choosing each element $i \in [n]$ randomly and independently with probability p . We show that the maximum possible size of (b/a) -multiple-free sets contained in $[n]_p$ is $\frac{b}{b+p}pn + O(\sqrt{pn} \log n \log \log n)$ with probability that goes to 1 as $n \rightarrow \infty$.

1. INTRODUCTION

In recent years a trend in extremal combinatorics concerned with investigating how classical extremal results in *dense* environments transfer to *sparse* settings, and it has seen to be a fruitful subject of research. Especially, in combinatorial number theory, the following extremal problem in a dense environment has been well-studied and successively extended to sparse settings: Fix an equation and estimate the maximum size of subsets of $[n] := \{1, 2, \dots, n\}$ containing no non-trivial solutions of the given equation.

An example of this line of research is a version of Roth's theorem [10] on arithmetic progressions of length 3 (with respect to the equation $x_1 + x_3 = 2x_2$) for random subsets of integers in Kohayakawa–Łuczak–Rödl [8]. Also, Szemerédi's theorem [12] was transferred to random subsets of integers in Conlon–Gowers [2] and Schacht [11]. The result of Erdős–Turán [4], Chowla [1], and Erdős [3] in 1940s on the maximum size of Sidon sets in $[n]$ was extended in [6, 7] to sparse random subsets of $[n]$, where a *Sidon*

Date: Thu 27th Dec, 2012, 3:04am.

The author was supported by Korea Institute for Advanced Study (KIAS) grant funded by the Korea government (MEST).

set is a set of positive integers not containing any non-trivial solution of $x_1 + x_2 = y_1 + y_2$.

In this paper we transfer the following extremal results to sparse random subsets. For a rational number $r > 1$, a set A of positive integers is called an r -multiple-free set if A does not contain any solution of $rx = y$. An interesting problem on r -multiple-free sets is of estimating the maximum possible size $f_r(n)$ of r -multiple-free sets contained in $[n] := \{1, 2, \dots, n\}$. This extremal problem has been studied in [14, 9, 13] for its own interest in combinatorial number theory, and also was applied to coding theory in [5].

Wang [14] showed that $f_2(n) = \frac{2}{3}n + O(\log n)$. Leung and Wei [9] proved that for every integer $r > 1$, $f_r(n) = \frac{r}{r+1}n + O(\log n)$. Wakeham and Wood [13] extended it to rational numbers as follows. For positive integers a and b , let $\gcd(b, a)$ be the greatest common divisor of a and b .

Theorem 1 (Wakeham and Wood [13]). *Let a, b be positive integers with $a < b$ and $\gcd(b, a) = 1$. Then*

$$f_{b/a}(n) = \frac{b}{b+1}n + O(\log n).$$

We shall investigate the maximum size of constant-multiple-free sets contained in a random subset of $[n]$. Let $[n]_p$ be a random subset of $[n]$ obtained by choosing each element in $[n]$ independently with probability p . Let $f_r([n]_p)$ denote the maximum size of r -multiple-free sets contained in $[n]_p$. We are interested in the behavior of $f_r([n]_p)$ for every rational number $r > 1$.

Theorem 1 gives the answer of the above question for the case $p = 1$. On the other hand, if $p = o(1)$, then the usual deletion methods give that *with high probability* (that is, with probability that goes to 1 as $n \rightarrow \infty$) the maximum size of (b/a) -multiple-free sets contained in $[n]_p$ is $np(1 - o(1))$. Hence, from now on, we consider p as a real number with $0 < p < 1$.

Using Chernoff bounds (for example, see Lemma 11), Theorem 1 easily implies the following:

Fact 2. *Let $p \in (0, 1)$ and let a, b be positive integers such that $a < b$ and $\gcd(a, b) = 1$. Let ω be a function of n that goes to ∞ arbitrarily slowly as $n \rightarrow \infty$. With high probability, there is a (b/a) -multiple-free set in $[n]_p$ of size*

$$\frac{b}{b+1}pn + \omega\sqrt{pn}.$$

Fact 2 gives a lower bound on $f_{b/a}([n]_p)$ that is off from the right value of $f_{b/a}([n]_p)$. The main result of this paper is the following:

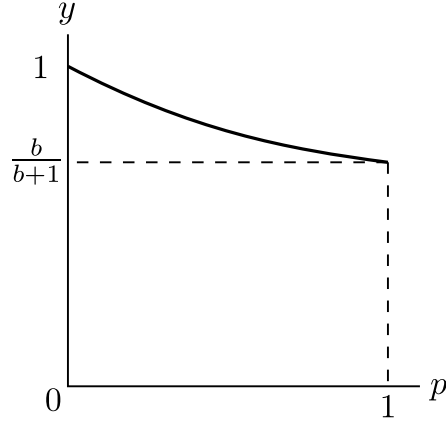


FIGURE 1. The graph of $y = b/(b+p)$ for $0 \leq p \leq 1$

Theorem 3. *Let $p \in (0, 1)$ and let a, b be positive integers such that $a < b$ and $\gcd(a, b) = 1$. Then, with high probability,*

$$f_{b/a}([n]_p) = \frac{b}{b+p}pn + O(\sqrt{pn} \log n \log \log n).$$

The ratio $\frac{f_{b/a}([n]_p)}{np}$ goes from 1 to $\frac{b}{b+1}$ as p varies from 0 to 1 (See Figure 1). The proof of Theorem 3 is given in Sections 2 and 3 by using a graph theoretic method.

2. PROOF OF THEOREM 3

In order to show Theorem 3, we use a graph theoretic approach that was used in Wakeham and Wood [13]. Let $r = b/a > 1$ be a rational number. Let $D = (V, E)$ be the directed graph with the vertex set $V = [n]$ in which the set E of arcs (or directed edges) is $\{(x, y) : rx = y\}$. Let $D[[n]_p]$ be the subgraph of D induced on $[n]_p$. Observe that $f_r([n]_p)$ is the same as the independence number $\alpha(D[[n]_p])$ of $D[[n]_p]$.

We consider structures of $D[[n]_p]$. The indegree and outdegree of each vertex in D are at most 1. Also, there is no directed cycle in D because $(x, y) \in E$ implies $x < y$. Therefore, each component of D or $D[[n]_p]$ is a directed path.

In order to obtain an independent set of $D[[n]_p]$ of maximum size, we consider such an independent set componentwise. Let C be a component of $D[[n]_p]$. As we mentioned above, C is a directed path. Let $V(C) = \{u_0, u_1, u_2, \dots, u_i, \dots, u_l\}$ be the vertex set of C such that $u_j < u_{j+1}$ and $(u_j, u_{j+1}) \in E$ for $0 \leq j \leq l-1$. Observe that $V^*(C) := \{u_0, u_2, u_4, \dots\} \subset$

$V(C)$ forms an independent set of C of maximum size. Therefore, the set

$$T^* := \bigcup_C V^*(C),$$

where C is each component of $D[[n]_p]$, forms an independent set of $D[[n]_p]$ of maximum size. Hence, we have the following.

Lemma 4. $f_r([n]_p) = |T^*|$.

Thus, in order to show Theorem 3, it suffices to show the following.

Lemma 5. *Let $p \in (0, 1)$ and let a, b be natural numbers such that $a < b$ and $\gcd(a, b) = 1$. Then, with high probability,*

$$|T^*| = \frac{b}{b+p}pn + O(\sqrt{pn} \log n \log \log n).$$

The proof of Lemma 5 is given in Section 3.

3. PROOF OF LEMMA 5

From now on, we show Lemma 5. For positive integers b and k , let k be an i -th subpower of b if $k = b^i l$ for some $l \not\equiv 0 \pmod{b}$. Let T_i be the set of i -th subpowers of b in $[n]$. Let $T_i^* \subset T_i$ denote the set of i -th subpowers v of b in $[n]_p$ such that v is at an even distance from the smallest vertex of the component of $D[[n]_p]$ containing v . Observe that $T^* = \bigsqcup_i T_i^*$, and hence,

$$|T^*| = \sum_i |T_i^*|. \quad (1)$$

In Section 3.1, we estimate the expected value $\mathbb{E}(|T^*|)$. Section 3.2 deals with a concentration result of $|T^*|$ with high probability.

3.1. Expectation. We first estimate $\mathbb{E}(|T_i^*|)$ and their sum $\mathbb{E}(|T^*|)$. Recall that T_i denotes the set of i -th subpowers of b in $[n]$. Note that since $1 \leq b^i \leq n$, the range of i is $0 \leq i \leq \log_b n$. It is clear that

$$T_i = \left\{ b^i x \mid 1 \leq x \leq \frac{n}{b^i}, \quad x \not\equiv 0 \pmod{b} \right\}.$$

Hence we have the following:

Fact 6.

$$|T_i| = \frac{b-1}{b} \frac{n}{b^i} \pm 1. \quad (2)$$

We consider two cases separately, based on the parity of i .

Lemma 7. *For $0 \leq j \leq (\log_b n)/2$, we have*

$$\mathbb{E}(|T_{2j}^*|) = \frac{b-1}{b(1+p)}pn \left(\frac{1}{b^{2j}} + \left(\frac{p}{b}\right)^{2j} p \right) \pm 1.$$

Proof. First we consider $\Pr [v \in T_{2j} \text{ is in } T_{2j}^*]$. Let $\{v_0, v_1, v_2, \dots\}$, where $v_i < v_{i+1}$, be the vertex set of the component of D containing v . Observe that $v_i \in T_i$, and hence, $v = v_{2j}$. The event that $v \in T_{2j}$ is in T_{2j}^* happens only when one of the following holds:

- There is some r with $0 \leq r \leq j-1$ such that $v_{2j-1-2r} \notin [n]_p$ and $v_i \in [n]_p$ for all $2j-2r \leq i \leq 2j$.
- The vertices v_0, v_1, \dots, v_{2j} are in $[n]_p$.

Hence, we have

$$\Pr [v \in T_{2j} \text{ is in } T_{2j}^*] = p((1-p) + p^2(1-p) + \dots + p^{2j-2}(1-p) + p^{2j}). \quad (3)$$

Thus we infer

$$\begin{aligned} \mathbb{E}(|T_{2j}^*|) &= |T_{2j}| \cdot \Pr [v \in T_{2j} \text{ is in } T_{2j}^*] \\ &\stackrel{(2),(3)}{=} \left(\frac{b-1}{b} \frac{n}{b^{2j}} \pm 1 \right) p \left((1-p) \frac{1-p^{2j}}{1-p^2} + p^{2j} \right) \\ &= \frac{b-1}{b(1+p)} pn \left(\frac{1}{b^{2j}} + \frac{p^{2j}}{b^{2j}} p \right) \pm 1, \end{aligned}$$

which completes the proof of Lemma 7. \square

Lemma 8. For $1 \leq j \leq (\log_b n)/2$, we have

$$\mathbb{E}(|T_{2j-1}^*|) = \frac{b-1}{b(1+p)} pn \left(\frac{1}{b^{2j-1}} - \left(\frac{p}{b} \right)^{2j-1} p \right) \pm 1.$$

Proof. Using an argument similar to the proof of (3), one may obtain that

$$\Pr [v \in T_{2j-1} \text{ is in } T_{2j-1}^*] = p((1-p) + p^2(1-p) + \dots + p^{2j-2}(1-p)). \quad (4)$$

Thus we infer

$$\begin{aligned} \mathbb{E}(|T_{2j-1}^*|) &= |T_{2j-1}| \cdot \Pr [v \in T_{2j-1} \text{ is in } T_{2j-1}^*] \\ &\stackrel{(2),(4)}{=} \left((b-1) \frac{n}{b^{2j}} \pm 1 \right) p(1-p) \frac{1-p^{2j}}{1-p^2} \\ &= \frac{b-1}{1+p} pn \left(\frac{1}{b^{2j}} - \left(\frac{p}{b} \right)^{2j} \right) \pm 1, \end{aligned}$$

which completes the proof of Lemma 8. \square

Lemmas 7 and 8 immediately imply the following.

Corollary 9. For $0 \leq i \leq \log_b n$, we have

$$\mathbb{E}(|T_i^*|) = \frac{b-1}{b(1+p)} pn \left(\frac{1}{b^i} + \left(-\frac{p}{b} \right)^i p \right) \pm 1. \quad (5)$$

Summing over all i with $0 \leq i \leq \log_b n$, we have the following.

Corollary 10.

$$\mathbb{E}(|T^*|) = \sum_{i=0}^{\log_b n} \mathbb{E}(|T_i^*|) = \frac{b}{b+p}pn + O(\log n).$$

Proof. One may easily see that for $|x| \geq b \geq 2$,

$$\sum_{i=0}^{\log_b n} \frac{1}{x^i} = \frac{x}{x-1} + O\left(\frac{1}{n}\right). \quad (6)$$

Corollary 9 yields that for $b \geq 2$

$$\begin{aligned} \sum_{i=0}^{\log_b n} \mathbb{E}(|T_j^*|) &\stackrel{(5)}{=} \sum_{i=0}^{\log_b n} \left[\frac{b-1}{b(1+p)}pn \left(\frac{1}{b^i} + \left(-\frac{p}{b}\right)^i p \right) \pm 1 \right] \\ &\stackrel{(6)}{=} \frac{b-1}{b(1+p)}pn \left[\frac{b}{b-1} + O\left(\frac{1}{n}\right) + \frac{-b/p}{-b/p-1}p + O\left(\frac{1}{n}\right) \right] + O(\log n) \\ &= \frac{b}{b+p}pn + O(\log n), \end{aligned} \quad (7)$$

which completes the proof of Corollary 10. \square

3.2. Concentration. Next we consider a concentration result of $|T_i^*|$. In other words, we show that $|T_i^*|$ is around its expectation with high probability. We will apply the following version of Chernoff bounds.

Lemma 11 (Chernoff bound). *Let X_i be independent random variables such that $\Pr[X_i = 1] = p_i$ and $\Pr[X_i = 0] = 1 - p_i$, and let $X = \sum_{i=1}^n X_i$. Then for any $\lambda \geq 0$,*

$$\Pr[X \geq (1 + \lambda)\mathbb{E}(X)] \leq e^{-\frac{\lambda^2}{2+\lambda}\mathbb{E}(X)}, \quad (8)$$

$$\Pr[X \leq (1 - \lambda)\mathbb{E}(X)] \leq e^{-\frac{\lambda^2}{2}\mathbb{E}(X)}. \quad (9)$$

In particular, for $0 \leq \lambda \leq 1$,

$$\Pr[|X - \mathbb{E}(X)| \geq \lambda\mathbb{E}(X)] \leq 2e^{-\frac{\lambda^2}{3}\mathbb{E}(X)}. \quad (10)$$

We first consider the case when $0 \leq i \leq 0.9 \log_b n$.

Lemma 12. *For $0 \leq i \leq 0.9 \log_b n$, we have*

$$|T_i^*| = \mathbb{E}(|T_i^*|) + O(\sqrt{pn} \log \log n) \quad (11)$$

with probability at least $1 - 2e^{-\frac{1}{3}(\log \log n)^2}$.

Proof. Fix i . If $k \in T_i \subset [n]$, then let

$$X_k = \begin{cases} 1 & \text{with probability } p^* \\ 0 & \text{with probability } 1 - p^*. \end{cases}$$

where $p^* = \Pr[v \in T_i \text{ is in } T_i^*]$. Otherwise, let $X_k = 0$ with probability 1. Let $X = \sum_{k=1}^n X_k$. Observe that

$$X = |T_i^*| \quad (12)$$

as random variables.

Note that for each $k \in T_i$, the event that $k \in T_i^*$ depends only on the events that $v \in [n]_p$, where the vertices v are in the component of D containing k and $v \leq k$. Hence, X_k are independent for all $k \in T_i$. Therefore we are able to use Chernoff bounds (Lemma 11) for a concentration result of X .

Set $\lambda = \frac{\log \log n}{\sqrt{\mathbb{E}(X)}}$. Note that $0 \leq \lambda \leq 1$ for $0 \leq i \leq 0.9 \log_b n$ since

$$\mathbb{E}(X) \geq \Omega\left(pn \frac{\varepsilon_p}{b^i}\right) \geq \Omega\left(pn \frac{\varepsilon_p}{n^{0.9}}\right) = \Omega(\varepsilon_p pn^{0.1}),$$

where ε_p is a positive constant such that $\varepsilon_p \rightarrow 0$ as $p \rightarrow 1$. The inequality (10) yields that

$$\Pr\left[|X - \mathbb{E}(X)| \geq \sqrt{\mathbb{E}(X)} \log \log n\right] \leq 2e^{-\frac{1}{3}(\log \log n)^2}. \quad (13)$$

Corollary 9 yields that $\mathbb{E}(|X|) = O(pn)$, and hence, we infer that

$$X = \mathbb{E}(X) + O(\sqrt{pn} \log \log n)$$

with probability at least $1 - 2e^{-\frac{1}{3}(\log \log n)^2}$. This together with (12) completes the proof of Lemma 12. \square

Next we consider the remaining case when $0.9 \log_b n \leq i \leq \log_b n$.

Lemma 13. *For $0.9 \log_b n \leq i \leq \log_b n$, we have $|T_i^*| = O((pn)^{0.1})$, with probability at least $1 - e^{-(\log \log n)^2}$.*

Proof. We define a random variable X as in (12), that is, $X = |T_i^*|$. Set $\lambda = \frac{2(\log \log n)^2}{\mathbb{E}(X)}$. The inequality (8) yields that

$$\Pr[X \geq (1 + \lambda)\mathbb{E}(X)] \leq e^{-\frac{\lambda}{2}\mathbb{E}(X)} = e^{-(\log \log n)^2},$$

and hence,

$$\Pr[X \geq \mathbb{E}(X) + 2(\log \log n)^2] \leq e^{-(\log \log n)^2}. \quad (14)$$

In other words,

$$X \leq \mathbb{E}(X) + 2(\log \log n)^2 \quad (15)$$

with probability at least $1 - e^{-(\log \log n)^2}$.

Corollary 9 gives that

$$\mathbb{E}(X) = O\left(pn \frac{1}{b^i}\right) = O(pn^{0.1}) = O((pn)^{0.1}), \quad (16)$$

where the second inequality holds for $i \geq 0.9 \log_b n$. Thus, combining (15) and (16) completes the proof of Lemma 13. \square

Now we are ready to show Lemma 5.

Proof of Lemma 5. We have that

$$|T^*| = \sum_{i=1}^{\log_b n} |T_i^*| = \sum_{i=1}^{\lfloor 0.9 \log_b n \rfloor} |T_i^*| + \sum_{i=\lfloor 0.9 \log_b n \rfloor + 1}^{\log_b n} |T_i^*|.$$

Lemmas 12 and 13 give that

$$|T^*| = \sum_{i=1}^{\log_b n} \mathbb{E}(|T_i^*|) + O(\sqrt{pn} \log n \log \log n),$$

with probability at least

$$1 - (\log_b n) \cdot 2e^{-\frac{1}{3}(\log \log n)^2} = 1 - 2e^{\log \log_b n - \frac{1}{3}(\log \log n)^2} = 1 - o(1). \quad (17)$$

This together with Corollary 10 implies that with high probability

$$|T^*| = \frac{b}{b+p}pn + O(\sqrt{pn} \log n \log \log n),$$

which completes the proof of Lemma 5. \square

Acknowledgement. *The author thanks Yoshiharu Kohayakawa for his helpful comments and suggestions, and thanks Jaigyoung Choe for his support at Korea Institute for Advanced Study.*

REFERENCES

- [1] S. Chowla. Solution of a problem of Erdős and Turán in additive-number theory. *Proc. Nat. Acad. Sci. India. Sect. A.*, 14:1–2, 1944.
- [2] D. Conlon and W. T. Gowers. Combinatorial theorems in sparse random sets. submitted, 70pp, 2010.
- [3] P. Erdős. On a problem of Sidon in additive number theory and on some related problems. Addendum. *J. London Math. Soc.*, 19:208, 1944.
- [4] P. Erdős and P. Turán. On a problem of Sidon in additive number theory, and on some related problems. *J. London Math. Soc.*, 16:212–215, 1941.
- [5] M. Jimbo, M. Mishima, S. Janiszewski, A. Y. Teymorian, and V. D. Tonchev. On conflict-avoiding codes of length $n = 4m$ for three active users. *IEEE Trans. Inform. Theory*, 53(8):2732–2742, 2007.
- [6] Y. Kohayakawa, S. Lee, and V. Rödl. The maximum size of a Sidon set contained in a sparse random set of integers. In *Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 159–171, Philadelphia, PA, 2011. SIAM.

- [7] Y. Kohayakawa, S. J. Lee, V. Rödl, and W. Samotij. The number of Sidon sets and the maximum size of sidon sets contained in a sparse random set of integers. Accepted to Random Structures & Algorithms.
- [8] Y. Kohayakawa, T. Łuczak, and V. Rödl. Arithmetic progressions of length three in subsets of a random set. *Acta Arith.*, 75(2):133–163, 1996.
- [9] J. Y.-T. Leung and W.-D. Wei. Maximal k -multiple-free sets of integers. *Ars Combin.*, 38:113–117, 1994.
- [10] K. F. Roth. On certain sets of integers. *J. London Math. Soc.*, 28:104–109, 1953.
- [11] M. Schacht. Extremal results for random discrete structures. submitted, 27pp, 2009.
- [12] E. Szemerédi. On sets of integers containing no k elements in arithmetic progression. *Acta Arith.*, 27:199–245, 1975. Collection of articles in memory of Juriĭ Vladimirovič Linnik.
- [13] D. Wakeham and D. Wood. On multiplicative Sidon sets. arXiv:1107.1073.
- [14] E. T. H. Wang. On double-free sets of integers. *Ars Combin.*, 28:97–100, 1989.

SCHOOL OF MATHEMATICS, KOREA INSTITUTE FOR ADVANCED STUDY (KIAS), SEOUL
130-722, SOUTH KOREA

E-mail address: sjlee242@gmail.com